

# Comparative analysis of Gaussian Process (GP) power curve models based on different stationary covariance functions for the purpose of improving model accuracy

Ravi Kumar Pandit, David Infield

University of Strathclyde, Glasgow

United Kingdom (UK)

ravi.pandit@strath.ac.uk / david.infield@strath.ac.uk

## Abstract

Gaussian Process (GP) models are increasingly finding application in wind turbine condition monitoring and in particular early fault detection. GP model accuracy is greatly influenced by the choice and type of the covariance functions (used to describe the similarity between two given data points). Hence, the appropriate selection and composition of covariance functions is essential for accurate GP modelling.

In this study, an in-depth analysis of commonly used stationary covariance functions is presented in which wind turbine power curve used where GP based power curve has been constructed using different stationary covariance functions, and after that, a comparative analysis has been carried out in order to identify the most effective covariance function. The commonly used squared exponential covariance function is taken as the benchmark, against which other covariance functions are assessed.

The results show that the performance (in terms of model accuracy and uncertainty) of GP fitted power curve models based on rational quadratic covariance functions is almost the same as for the most commonly used squared exponential function. Thus, rational quadratic covariance functions can be used instead of squared exponential covariance functions. In this paper, strength and weakness of stationary covariance functions would be highlighted for effective condition monitoring.

**Keywords:** condition monitoring, power curve, covariance functions, Gaussian Process.

## List of List of abbreviations

|          |  |
|----------|--|
| ECov     | Exponential covariance function          |
| SECOv    | Square Exponential covariance function   |
| $k_{RQ}$ | Rational quadratic covariance function   |
| COE      | Cost of Energy                           |
| CIs      | Confidence Intervals                     |
| GP       | Gaussian Process                         |
| MAE      | Mean Absolute Error                      |
| O&M      | Operation & Maintenance                  |
| RMSE     | Root Mean Square Error                   |
| RQ       | Rational quadratic                       |
| SCADA    | Supervisory Control And Data Acquisition |
| VSWTGs   | Variable speed wind turbine generators   |
| WTs      | Wind Turbines                            |

## 1. Introduction

Power generation from the wind has been making steady progress in recent years. A recent study [1] found that the total global installed wind capacity reached 486 GW by the end of 2016 which represents an annual growth rate of 11.8 %, where China's contribution is presently the highest. Reducing the Cost of Energy (COE) has been crucial to this success. Increasing the availability of WTs and reducing their operation and maintenance (O&M) costs are attractive measures to reduce the COE further.

Condition based maintenance is expected to reduce O&M costs, especially offshore, but must be underpinned by effective and low-cost condition monitoring (CM). As of now, significant research and innovation effort is being applied to condition monitoring systems (CMS) development and technology advancement [2,3]. Remote areas are commonly used for both onshore and offshore wind deployments due to their richer wind resource and the advantages of mitigating the land use and visual impact issues. Specifically, for offshore wind farms, site access difficulties and the shortage of proper transportation, logistics and installation vehicles is making O&M more challenging and costly [4]. Developing SCADA based condition monitoring is a cost-effective approach to reducing the COE.

Due to fluctuations in wind speed and power, precise performance assessment of a wind turbine can be challenging. An ageing turbine may be subject to faults and component failures, including catastrophic failures. Unscheduled maintenance due to unexpected failures of wind turbine components is considered to be a dominant reason for high maintenance costs. However, developing an efficient automated fault detection algorithm for continuous monitoring of wind turbines can help improve maintenance so increase reliability which ultimately reduces the cost of electricity. In past, [5, 6, 7, 8], various nonparametric models proposed to assess turbine health in terms of power performance curves. The power curve quantifies power production as a function of hub height wind speed. With the help of the power curve, performance deterioration associated with a range of faults can be identified.

A Gaussian Process (GP) is a nonparametric approach based on a general stochastic model. Due to the simplicity and ease of construction, GP models are widely used in forecasting and or fitting related applications. For example, in [9], it is suggested that GP based model for power forecasting provides 4% to 11% improvement in prediction accuracy over an artificial neural network (ANN) model (another nonparametric approach). In another article, [10], GP models able to detect the anomaly due to misalignment without any false alarms. The constructed GP algorithm is then compared with two fault algorithms based on binning, namely, Probabilistic assessment of incoming data using a real-time power curve and Probabilistic assessment of incoming data using a binned power curve. The comparative analysis suggests that, GP algorithm able to detect failure 1.5 hours after the first sign of failure while two probabilistic binning method took 4 hours and 6 hours respectively.

## **Scientific contributions**

A GP provides a natural and straightforward way to estimate the uncertainty associated with its predictions due to its intrinsic confidence intervals. The confidence intervals come with GP estimated values and can play an important role in assessing the performance of wind turbines. For example, in [11] the present authors used the confidence interval of a GP power curve model to determine whether a turbine is underperforming or not in the specific case of wind turbine yaw misalignment. GP models are constructed from mean and covariance functions. The choice of covariance function is a critical factor for accurate GP model development. There are various types of covariance functions available, see [12]. Frequently, squared exponential covariance (SECOV) functions are used for power curve fitting, for example [13], but there is limited research

investigating how GP models behave with other covariance functions in terms of fitting accuracy and uncertainty. The objective of this paper is to address this issue by investigating a range of popular stationary covariance functions covering different classes. These are Matern class (i.e. Matern 3/2, Matern 5/2); Exponential class (i.e. exponential, squared exponential) and rational quadratic covariance functions. Uncertainty associated with model estimation is central to dealing with fault detection, and so this research also examines the covariance function impact on model uncertainty. The advantages and disadvantages of using alternative covariance functions for GP wind turbine power curve fitting will be determined. The power curve is used to describe the nonlinear relationship between wind speed and power output from a wind turbine and is widely used by wind engineers to assess the power performance, wind turbine selection, condition monitoring, wind energy assessment and predictive maintenance. For these reasons power curve accuracy is used as key indicator of the impact of different covariance functions on GP model effectiveness.

## **2. Wind Turbine SCADA data description and pre-processing**

Wind turbine condition monitoring systems can be expensive, and not all operators regard them as cost-effective. However, with the increase in age of wind turbines and the move to less accessible offshore sites, the O&M cost is expected to increase which reinforces the drive towards condition based maintenance. A Supervisory Control and Data Acquisition (SCADA) system collects signal that contains large amounts of raw and useful information from critical wind turbine sub-assemblies. Extract key information from the SCADA data is an economical and effective way for condition assessment since the data available is free and does not require any extra sensors. SCADA provides crucial information about load history and the operation of individual turbines and thus can be helpful in identifying early warning of failures.

The SCADA datasets used in this paper are from operational wind turbines located in Scotland, UK, and contain more than 100 different signals, ranging from the timestamp, calculated values, set point, measurements of temperature, current, voltage, wind speed, power output, wind direction and so on. All the SCADA data consists of 10-minute averages with maximum, minimum, standard deviation over the 10 minutes also being recorded and correspond to a full year of operation. The turbines used in this paper are from Siemens 2.3 MW turbines. The SCADA datasets for 2012 have been used for GP model testing and validation.

SCADA data is not without errors due to sensor failure and data collection faults. Such errors will affect power curve accuracy and should be systematically removed at the outset. Criteria like timestamp mismatches, out of range values, negative power values, and turbine power curtailment are considered to remove misleading data as per methodologies used in [14]. It should be noted; these adopted methodologies would not make SCADA data completely error free but minimise these errors. Table 1 is the summary of SCADA dataset with start time stamp “1/1/2012 00:00 AM” and end timestamp “31/1/2012 23:50 PM” of total measured observation is equal to 4464, and after filtrations, the data point came down to 1926. Moreover, this filtered data is used for GP model training and validation purposes in upcoming sections.

| Dataset | Start timestamp   | End timestamp      | Measured data | Filtered data |
|---------|-------------------|--------------------|---------------|---------------|
| 1       | 1/1/2012 00:00 AM | 31/1/2012 23:50 PM | 4464          | 1926          |

Table 1: SCADA dataset description

Figure 1 shows the measured power curve of an individual wind turbine before data filtering, and Figure 2 shows the power curve after filtering as outlined above. Note that in Figure 2 air density corrections following the IEC recommendations, as described in the next section, have been implemented.

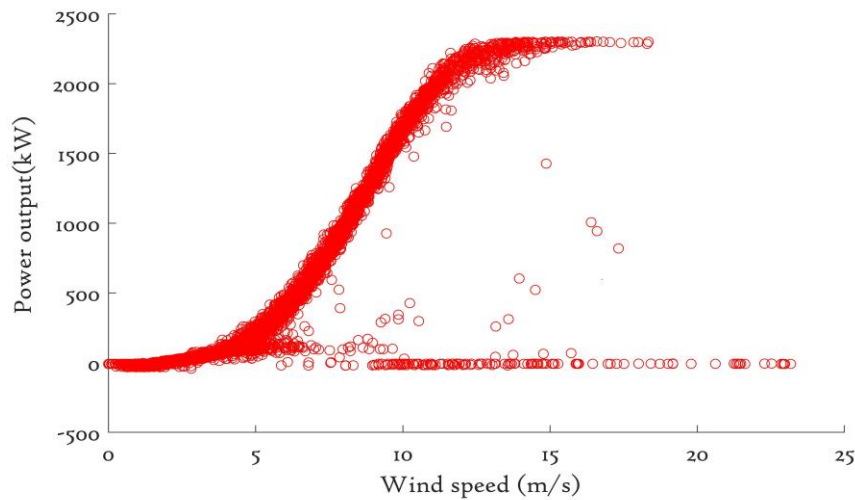


Figure 1: Scatter plot of measured power curve

### 3. Wind turbine power curves and air density correction

Variable speed wind turbine generators (VSWTs) wind turbines are currently most widely used and the power curve equation, as given below, is used for performance evaluation and monitoring purposes, see for example [15,16]. Moreover, wind power assessment of a region is based on hub height wind speed, and air density, with the wind power potential, evaluated using a power curve, [17]. The relationship between the power output and wind speed is nonlinear in nature and describe by following mathematical equation,

$$P = 0.5 \rho A C_p(\lambda, \beta) v^3 \quad (1)$$

where  $\rho$  is Air density ( $kg/m^3$ ),  $A$  is swept area ( $m^2$ ),  $C_p$  is power coefficient of wind turbine and  $v$  is the hub height wind speed ( $m/sec$ ).

The equation (1) signifies the actual power production but depends critically on the power coefficient,  $C_p$ . The power coefficient itself is the function of tip speed ratio ( $\lambda$ ) and blade pitch angle ( $\beta$ ). In addition, the power capture of a wind turbine depends on other parameters such as wind direction, wind shear and other turbine parameters [18].

The altitude, weather induced changes in air pressure, and ambient temperature at the wind site affect the air density. Therefore, in order to analyse the performance of modern variable speed turbines, an air density correction should be applied as per IEC standard 61400-12-1, [18] and is given as follows,

$$\rho = 1.225 \left[ \frac{288.15}{T} \right] \left[ \frac{B}{1013.3} \right] \quad (2)$$

and, 
$$V_C = V_M \left[ \frac{\rho}{1.225} \right]^{\frac{1}{3}} \quad (3)$$

where,  $V_C$  and  $V_M$  are the corrected and measured wind speed in m/sec and the corrected air density is calculated by equation (2) where  $B$  is atmospheric pressure in mbar and  $T$  the temperature in Kelvin. The 10-minute average values of  $B$  and  $T$  are obtained from SCADA datasets. The calculated value of  $\rho$  from equation (3) is used to calculate the corrected wind speed ( $V_C$ ). An interesting result, presented in [28] established that IEC prescribed air density correction may not give most accurate power curve and suggested that adding air density directly into the GP model without any pre-correction is the best air density correction approach. However, in the present

study the IEC air density correction methodology is adopted. The power curve shown in Figure 2 shows preprocessed data (based on criteria outlined in section 2) that has been corrected for air density. Note that a 500 kW power generation threshold has been applied in order to concentrate on operation with non-trivial power output.

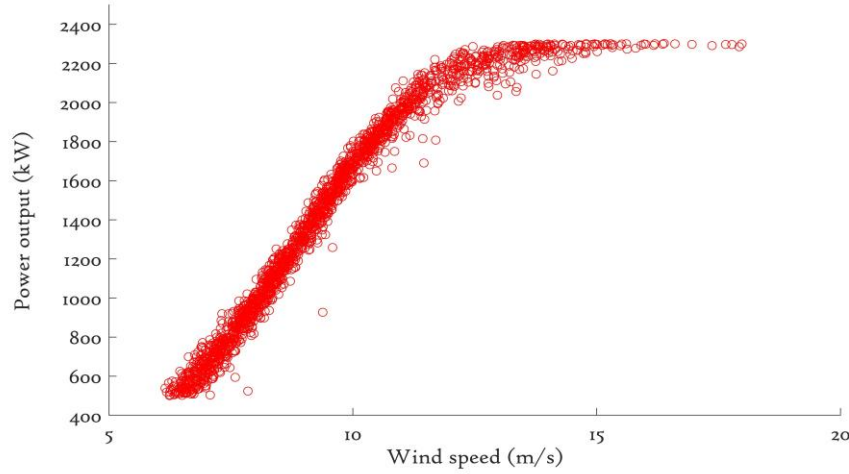


Figure 2: Scatter plot of pre-processed and air density corrected power curve

#### 4. Gaussian process models

A Gaussian Process is a multivariable probabilistic based generic supervised machine learning approach used to address problems related to regression and probabilistic classification.

GPs models generalise the multivariate Gaussian distribution from finite-dimensional vectors to infinite dimensionality. Notably, a GP is a stochastic process that has Gaussian-distributed finite-dimensional marginal distributions defined over a distribution of functions, i.e. each output from the GP is itself a function. GPs are widely used in probabilistic regression problems, for example, [12,19], thanks to their flexibility and simplicity of construction.

A GP, in essence, is a nonparametric generalisation of a joint normal distribution for a given potentially infinite set of variables, and it is mathematically defined by its mean and covariance functions (or kernel) as given in equation (4),

$$Y \sim GP(\mu, \Sigma) \quad (4)$$

where,  $\mu$  is the mean function, and  $\Sigma$  is the covariance function that has an associated probability density function:

$$P(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \quad (5)$$

Where  $|\Sigma|$  is defined as the determinant of  $\Sigma$ ,  $n$  is the dimension of random input vector  $x$ , and  $\mu$  is mean vector of vector  $x$ . The term under exponential i.e.  $\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)$  is an example of a quadratic form.

Gaussian Processes not only relax the assumption of linearity in the regression parameters but also give freedom where we do not need to specify a precise functional form for the relationship between predictor and response variables. The advantages of GP are summarised as follows [12,19].

### **Advantages**

- GP models, not only estimate predicted values but also come with intrinsic confidence intervals. These confidence intervals signify the uncertainty associated with the prediction and are very useful for anomaly detection which in the intended application are wind turbine faults.
- GP models are very versatile due to the abundant choices available for covariance function for any given problem
- With a GP, the predicted values are interpolations between the training data points that conform to the chosen covariance functions, so that the selection of covariance function controls the nature of the interpolation.

Uncertainty assessment is the key for developing a fault detection algorithm based on the GP. The uncertainty in the GP expressed by its confidence intervals whose robustness is directly related to covariance function, as will be discussed in upcoming sections.

## **5. Stationary covariance functions for Gaussian Process power curve fitting**

Covariance functions are commonly known as positive (semi-) definite kernels or Mercer's kernels, [20] and considered to be the driver of the fitting quality of the models along with hyperparameters (described in section 4). The covariance function is the decisive ingredient in Gaussian Process modelling since it encodes the assumptions about the function which we wish to learn. A covariance function describes the dependency of two variables concerning each other and



is the heart of the GP model; it signifies the similarity between two points and hence determines closeness between two points.

The reliability of GP regression is dependent on what and how well we select the covariance function. The covariance matrix,  $K$ , gives the variance of each variable along the leading diagonal, and the off-diagonal elements measure the correlations between the different variables, and are

given by: 
$$K = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{n1} & \cdots & k_{nn} \end{bmatrix} \quad \text{where } k_{ij} = k(x_i, x_j)$$

$K$  is of size  $n \times n$ , where  $n$  is the number of input parameters considered, and it must be symmetric and positive semidefinite i.e.  $\sum_{ij} = \sum_{ji}$ .

The basics properties of covariance functions are [12,20]:

- ✓ It must be positive semi-definite.
- ✓ It can be non-stationary (e.g., the length-scale may depend on the values of  $x$ ).
- ✓ It can be the sum (or product or linear combination) of other covariance functions, e.g., can use a different covariance function for each unique sensor modality or data type (vector, sequence, image data).

In general, covariance functions are classified into two categories; i) stationary and ii) nonstationary. A stationary covariance function is one that only depends on the relative position of its two inputs, and not on their absolute location in time. In short, stationary covariance function is invariant to translations in the input space. Exponential class, Matern class and rational quadratic are examples of stationary covariance functions while examples of nonstationary covariance functions are dot product and polynomial, [12].

Stationary covariance functions are the focus of this paper as these are appropriate to GP power curve model construction, described as follows.

### ***5.1 Exponential class covariance functions***

Two types of covariance functions fall into the category of exponential class; i) Exponential covariance function (ECov) and ii) Squared exponential covariance function (SECov).

- **Exponential covariance function (ECov)**

The exponential covariance function (ECov) for any finite collection of inputs  $\{x_1, x_2, \dots, x_n\}$  is defined as,

$$k_E(x, x') = \sigma_f \exp\left(-\frac{(x-x')}{l}\right) \quad (6)$$

where,  $(x - x')$  is the distance between  $x$  and  $x'$ ,  $\sigma_f$  is the signal variance and  $l$  is the characteristic length. In order to include the effect of measurement errors associated with the SCADA data, an additional noise variance  $\sigma_n^2$  is added to equation (6), which is rewritten as,

$$k_E(x, x') = \sigma_f \exp\left(-\frac{(x-x')}{l}\right) + \sigma_n^2 \delta(x, x') \quad (7)$$

Using equation 7, a power curve based on GP theory (described in section 6) is constructed, and is shown in Figure 3. Figure 3 indicates a clear case of overfitting and suggests that GP algorithm constructed using the exponential covariance function gives a fitting that tries to follow too much of the random variation in the data, and consequently is far from the desired smooth function that is expected for a wind turbine power curve. Note: the red line in the Figure represents the confidence intervals, discussed in the next section.

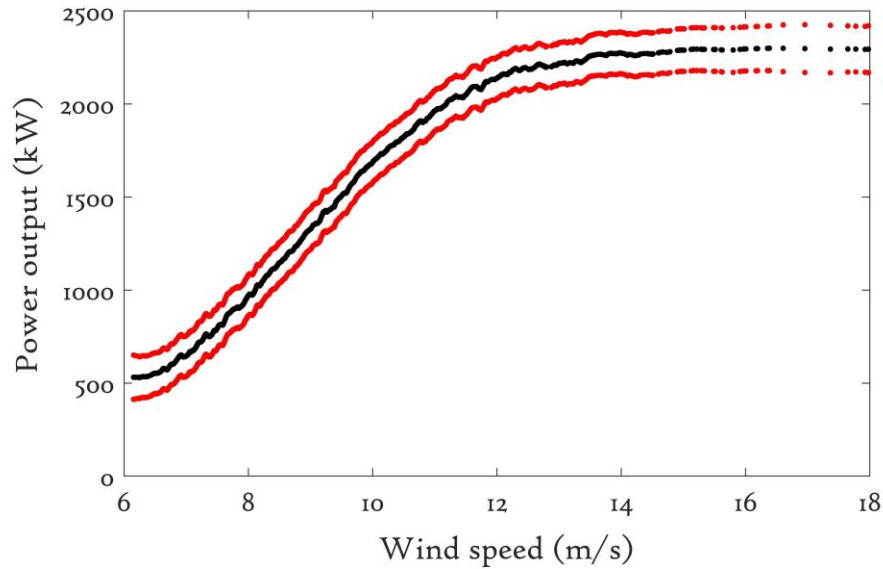


Figure 3: Power curve fitting using exponential covariance function

- **Squared exponential covariance function (SECov)**

The squared exponential covariance function (SECov) is a modified form of exponential covariance function and a widely used default covariance function in GP and SVM (Support vector machine) models, [12,20]. For any finite collection of inputs  $\{x_1, x_2, \dots, x_n\}$ , it is defined as:

$$k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l^2}\right) \quad (8)$$

$\sigma_f^2$ ,  $\sigma_n^2$  and  $l$  are known as the hyper-parameters and suitable optimization of these parameters is the key to accurate GP model fitting, [21].

Since the squared exponential covariance function popularly used and therefore, it will be used as the benchmark against which comparative studies of stationary covariance function is being carried out and assess in terms of the model smoothness and fitting accuracy.

### 5.2 Matern class covariance functions

The Matern class of covariance functions was discovered by the Swedish statistician Bertil Matérn and is widely used in spatial statistics, [22], machine learning, [12], geostatistical analysis, [23], and other applications involving multivariate statistical analysis on metric spaces. It is commonly used to define the statistical covariance between observations made at two points that are  $r$  units distant from each other. The Matern function depends solely on  $r$  and hence is stationary in nature. The distance used here is Euclidean distance, hence the Matérn covariance function is also isotropic. Mathematically it is defined as,

$$k_v(r) = \sigma_f^2 \frac{2^{1-v}}{\Gamma(v)} \left(\sqrt{2v} \frac{r}{l}\right)^v C_v\left(\sqrt{2v} \frac{r}{l}\right) \quad (9)$$

where  $\Gamma$  is the gamma function,  $C_v$  is the modified Bessel function of the second kind, [24], and  $l$  and  $v$  are non-negative parameters.

If  $v = l + \frac{1}{2}$ , then the Marten covariance can be expressed as a product of an exponential and a polynomial of order  $l$ , [25], as shown below,

$$k_{l+\frac{1}{2}}(r) = \sigma_f^2 \exp\left(-\frac{\sqrt{2v}r}{l}\right) \frac{\Gamma(l+1)}{\Gamma(2l+1)} \frac{2^{1-v}}{\Gamma(v)} \sum_{i=0}^l \frac{(l+i)!}{i!(l-i)!} \left(\frac{\sqrt{8vr}}{l}\right)^{l-i} \quad (10)$$

If  $v = \frac{3}{2}$  and  $l = 1$  then equation (11), after simplification reduces to

$$k_{3/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{l}\right) \exp\left(\frac{-\sqrt{3}r}{l}\right) \quad (11)$$

Moreover, if  $v = \frac{5}{2}$  and  $l = 2$  then equation (11), after simplification reduces to

$$k_{5/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(\frac{-\sqrt{3}r}{l}\right) \quad (12)$$

Where  $r = \sqrt{(x - x')^T (x - x')}$  is the Euclidean distance between  $x$  and  $x'$ .

Equations (11) and (12) are the mathematical expressions for Matern 3/2 and Matern 5/2 covariance functions respectively.

Particular case: when  $\nu \rightarrow \infty$ , then the Matern function converges to the squared exponential covariance function, [25], as shown below.

$$\lim_{\nu \rightarrow \infty} k_\nu(r) = \sigma_f^2 \exp\left(-\frac{r^2}{2l^2}\right) \quad (13)$$

Hence, the Matern covariance function includes the squared exponential covariance function as a special case.

### 5.3 Rational quadratic (RQ) covariance function

The mathematical equation for rational quadratic covariance functions for points  $x$  and  $x'$  is defined below, [26],

$$k_{RQ}(x, x') = \sigma_f^2 \left(1 + \frac{(x-x')^2}{2\alpha l^2}\right)^{-\alpha} \quad (14)$$

Where  $\alpha$  is a positive-valued scale-mixture parameter and determines the relative weighting of large-scale and small-scale variations. If we compare the rational quadratic function with the squared exponential covariance function, it is found that rational quadratic is equivalent to the adding a series of squared exponential covariance function with different values of the length scale,  $l$ . Thus, theoretically, it is expected that rational quadratic fits the data smoothly across different length scales.

SCADA data comes with measurement errors so, as mentioned above, it is desirable to add a noise term (i.e. calculated noise value of Table 1) to the covariance function in order to improve the accuracy of the GP model. Hence equations (8), (11), (12) and (14) are modified to be:

$$k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l^2}\right) + \sigma_n^2 \delta(x, x') \quad (15)$$

$$k_{3/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{l}\right) \exp\left(\frac{-\sqrt{3}r}{l}\right) + \sigma_n^2 \delta(x, x') \quad (16)$$

$$k_{5/2}(r) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(\frac{-\sqrt{3}r}{l}\right) + \sigma_n^2 \delta(x, x') \quad (17)$$

$$k_{RQ}(x, x') = \sigma_f^2 \left(1 + \frac{(x-x')^2}{2\alpha l^2}\right)^{-\alpha} + \sigma_n^2 \delta(x, x') \quad (18)$$

Here,  $\sigma_f^2$  signifies the signal variance and  $l$  is a characteristic length scale which describes how quickly the covariance decreases with the distance between points.  $\sigma_n$  is the standard deviation of the noise term and impacts on model uncertainty,  $\delta$  is the Kronecker delta, [12], and optimization of these hyper parameter was described in upcoming section.

## 6. Gaussian Process power curve modelling and hyperparameter optimization

The SCADA datasets contain measurements of a range of parameters from an operational wind turbine, but these are of course subject to measurement noise. Theoretically, the underlying function of the data being modelled is represented by following equation:  $y = f(x) + \epsilon$  where  $x$  are values from the training datasets and  $\epsilon$  is Gaussian white noise of variance  $\sigma_n^2$  such that  $\epsilon = N(0, \sigma_n^2)$ . Like many Bayesian modeling approaches, a prior distribution is vital for GP modeling, which represents the significant information about uncertain parameters. This together with the with the probability distribution of new incoming data is used to generate the posterior distribution. The obtained posterior distribution is significant for future inference and any decisions involving of uncertain parameters, for more detail see [27]. It is worth noting that the prior can be uninformative or informative. A Gaussian process prior is with the possible range of underlying functions of  $f(x)$  with unknown hyper-parameters and different stationary covariance functions described in section 5, hence the functions become

$$y_1, y_2, y_3, \dots \dots y_n \sim N(0, k) \quad (19)$$

$$\text{And,} \quad K = \Sigma + \sigma_n^2 I \quad (20)$$

where,  $\sigma_n^2 I$  is the product of variance and identity matrix ( $n \times n$ ) and describe the covariance between the output due to white noise and the training output data modelled using the following equation,

$$y_i = f(x_i) + \epsilon_i \quad (21)$$

Predicting the mean value ,and variance are the necessary estimation steps for GP algorithms. For the current scenario, our objective is to estimate the power  $y'$  for a new wind speed value  $x'$  for a training data say,  $D = \{(U_i, P_i), i = 1, \dots, N\}$ . The training datasets used to calculate the posterior distribution of  $P_*$  for a given input  $U_*$  and is defined as  $p(P_* | U_*, U_{tr}, P_{tr})$  in which  $\{P_*, U_*\}$

is the future power and wind speed values.  $P_{tr}, U_{tr}$  are the training SCADA datasets of power and wind speed.

For a given covariance function, the fitting is straightforward and involves work related to matrix manipulations. However, which covariance function best complements the problem is difficult to determine, hence detailed investigation of the sort undertaken here needs to be done. The covariance functions depend on the hyper-parameters which need to be optimized before the posterior distribution of  $P_*$  is calculated. In short, the covariance function depends on the hyper parameters and proper optimization of these ensure GP model accuracy. Take an example, the squared exponential covariance function (described in section 5) being used for power curve fitting; its hyper parameters ( $\sigma_f^2$ ,  $\sigma_n^2$  and  $l$ ) can be optimized via maximization of the log marginal likelihood. This is well described in [12,28] and is given below,

$$\log(p(P_{tr}|U_{tr})) = -0.5P_{tr}^TK^{-1}P_{tr} - 0.5\log(|K|) - 0.5n\log(2\pi) \quad (22)$$

Once the hyper parameters have been optimised using equation (22), the estimation of the distribution of  $P_*$  for a given  $U_*$  is straightforward. The estimated distribution of  $P_*$ ,  $p(P_*|U_*, U_{tr}, P_{tr})$  follows a Gaussian distribution with mean and variance as follows,

$$\mu(P_*) = k_*^TK^{-1}P_{tr} \quad (23)$$

$$\sigma^2(P_*) = k_{**} - k_*^TK^{-1}k_* + \sigma_n^2 \quad (24)$$

Where  $k_* = [k(U_*, U_1)k(U_*, U_2)k(U_*, U_3) \dots \dots k(U_*, U_n)]^T$  is the covariance's between test and training data points in the form of column vector and  $k_{**} = k(U_*, U_*)$  is the auto covariance function of the testing data points. The posterior variance of equation (24) depends on  $k_*$  hence it will be inversely proportional to the distance between test and training data points while the mean estimation of equation (23) is the linear combination of the output  $P_{tr}$  in which linear weights are defined  $k_*^TK^{-1}$ . A similar approach has been taken for power curve fitting using GP and hyper-parameters optimization equations for Matern class as described in [12,29] and rational quadratic covariance as described in [12,30]. The resulting values of optimized hyper-parameters for the different covariance functions are summarized in Table 2.

| <b>Hyperparameters</b>                 | <b>SECov</b> | <b>Matern<sub>3/2</sub></b> | <b>Matern<sub>5/2</sub></b> | <b>RQ</b> |
|--|--------------|-----------------------------|-----------------------------|-----------|
| length scale - $l$                     | 4.21         | 18.92                       | 8.61                        | 4.21      |
| signal standard deviation - $\sigma_f$ | 734.54       | 1364.51                     | 970.58                      | 734.63    |
| scale-mixture parameter - $\alpha$     | -            | -                           | -                           | 2985.51   |

Table 2: calculated hyperparameters of the GP power curve model for different stationary covariance function

The initial value for the noise standard deviation used to calculate GP models  $\sigma_n$  and is calculated by :  $\frac{std(y)}{\sqrt{2}}$  where y is the response variable, using MATLAB [12,31] ,shown in Table 3.

| <b>GP models</b>                            | <b>SECov</b> | <b>Matern<sub>3/2</sub></b> | <b>Matern<sub>5/2</sub></b> | <b>RQ</b> |
|---|--------------|-----------------------------|-----------------------------|-----------|
| <b>Initial noise(<math>\sigma_n</math>)</b> | 55.11        | 55.05                       | 55.10                       | 55.11     |

Table 3: Initial noise ( $\sigma_n$ ) of GP power curve models for different covariance functions

Using the filtered power curve SCADA data (Figure 2) and the covariance functions described in Section 5 together with the optimised hyperparameters, GP power curve models have been determined using MATLAB fitting procedures; these are shown in Figure 4 together with estimated 95% confidence intervals. Here, a QR factorisation approach was used for computing the log-likelihood and gradient for power curve estimation since this approach gives better accuracy as compared to the V-method-based approach, [12,31]. Figure 5 show power in time series for different types of GP models based on stationary covariance functions.

A quasi-Newton optimiser method was used to optimise the hyper-parameters concerning the likelihood in simple ML-II fashion, [32]. This optimiser approximates the Hessian and uses a trust-region method with a dense, symmetric rank-1-based(SR1). For other available optimisers see [12,24].

It is worth noting that irrespective of covariance function used in GP model, the computation of the marginal likelihood requires a matrix inversion and thus has asymptotic complexity called cubic inversion  $O(N)^3$  where  $N$  is the number of data points. If  $N$  contains large number of data points, then the computation of the  $N \times N$  matrix becomes problematic and leads to GP model inaccuracy. In order to solve this problem, methods based on state-space being described in

[33,34]. However, these methods need high processing power and there is a significant computational cost. Hence striking the balance between the data points and computation cost is the key for constructing an effective GP algorithm for the purpose of fault detection.

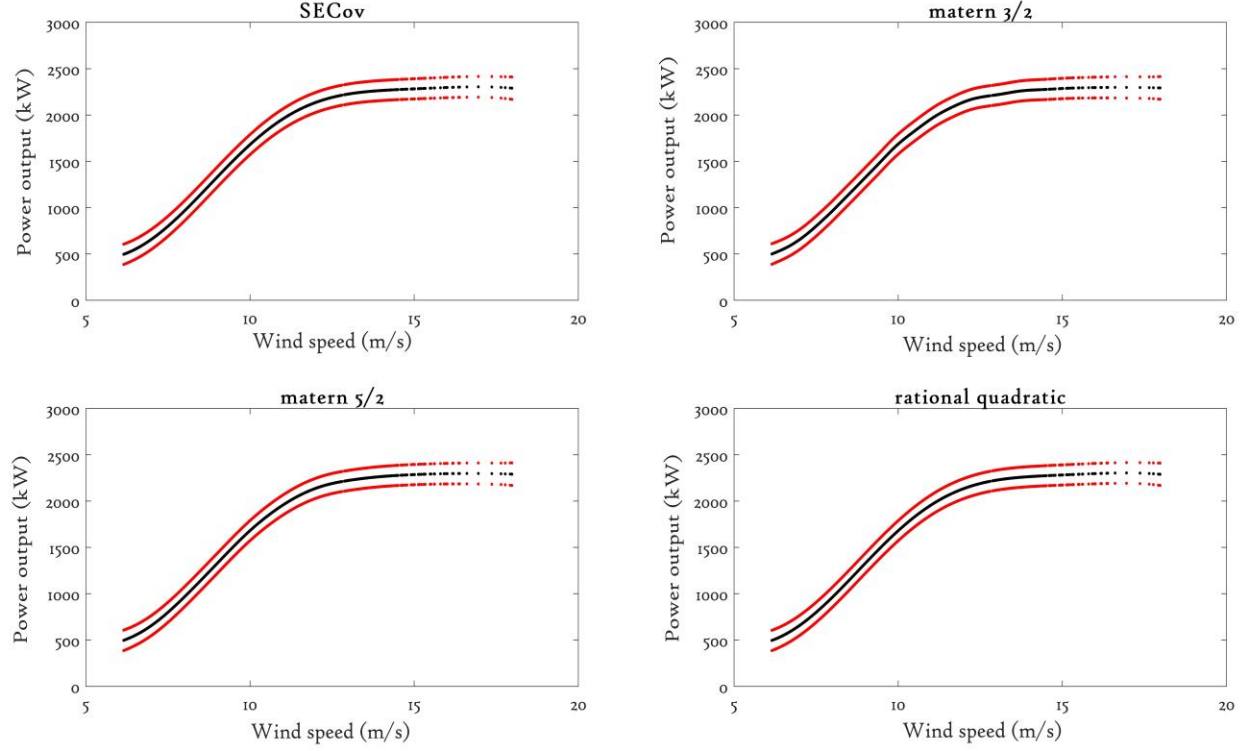


Figure 4: Modelled power curves with confidence intervals for different covariance functions

Due to the non-parametric, non-linear behaviour of a GP model, the residual which indicates the difference between measured and predicted values should be investigated. Theoretically, residuals of a GP model should be Gaussian, and this can be used to assess different stationary covariance function based GP power curve models for fitting accuracy and the nature of their distribution functions. The frequency distribution of the residuals is shown in Figure 6 together with a fitted Gaussian distribution; it can be seen that the rational quadratic and squared exponential covariance function based GP models have more or less identical distributions of residuals.



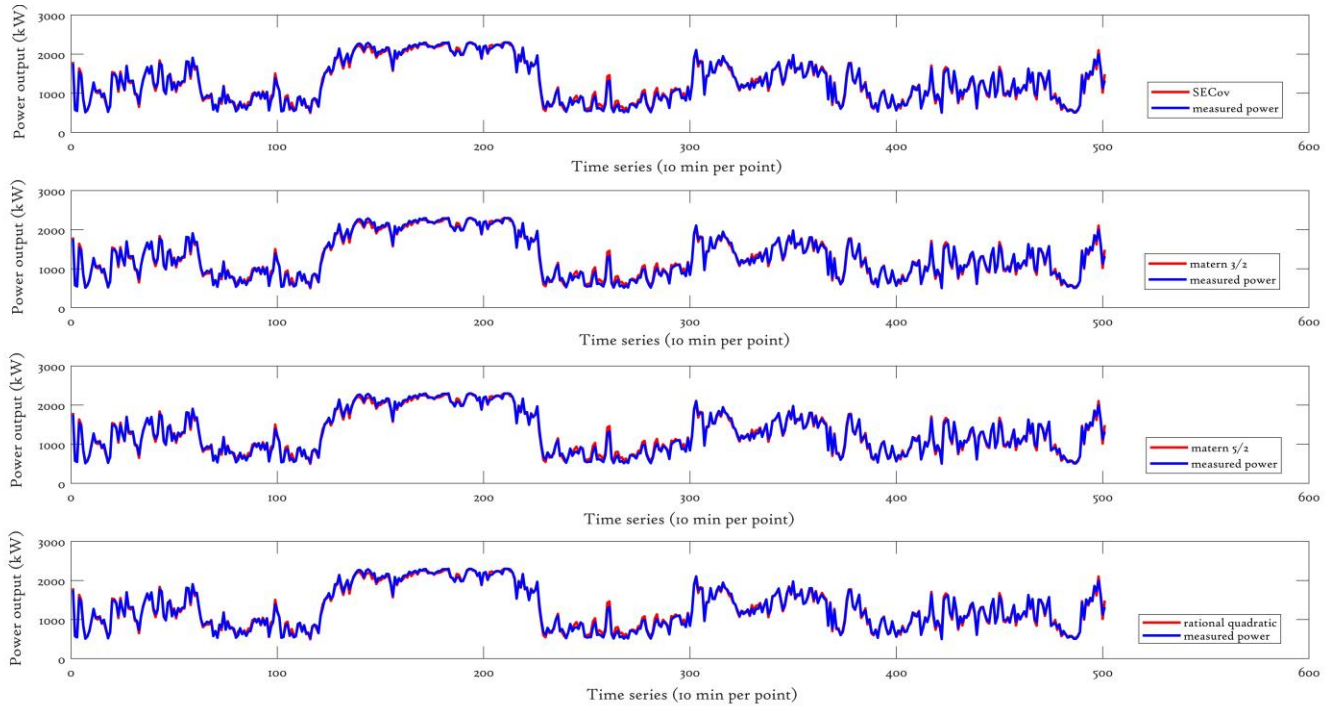


Figure 5: Estimated & measured power values

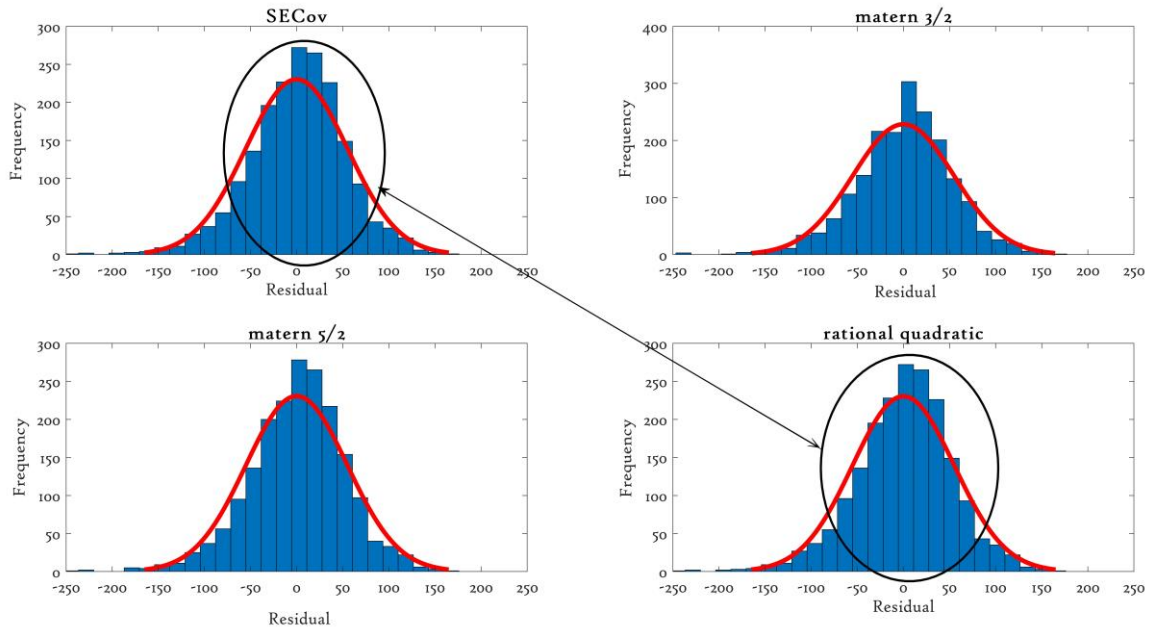


Figure 6: Histogram distribution fits for the different GP models

## 7. Gaussian Process Uncertainty analysis using confidence intervals

In order to assess GP power curve accuracy, confidence intervals can be used. The confidence intervals (CIs) are a useful measure of uncertainty and the precision of model estimates. Confidence intervals provide significant information about the uncertainty surrounding an estimation. The author of, [35], carried out the comparative analysis of uncertainty in power curve based on IEC binning method using error bars and GP using CIs and found out that the GP CIs are able to reflect the uncertainty around power curve.

As already mentioned, the GP model based on a squared exponential covariance function will be used as a benchmark in order to assess other GP models' accuracy. Figures 7,9 and 11 shows the comparisons of fitted GP models based on different covariance functions with the measured power curve, while Figures 8,10 and 12 compare model uncertainty as a function of the wind speed quantified in terms of confidence intervals against the benchmark. For the Matern 3/2 based GP model, the uncertainty is high, while with the Matern 5/2 based GP model it is small, but model fitting accuracy is less as compared to the benchmark GP model. The power curve fitting and uncertainty performance are almost the same for the benchmark model and GP model based on a rational quadratic covariance function, as shown in Figure 12.

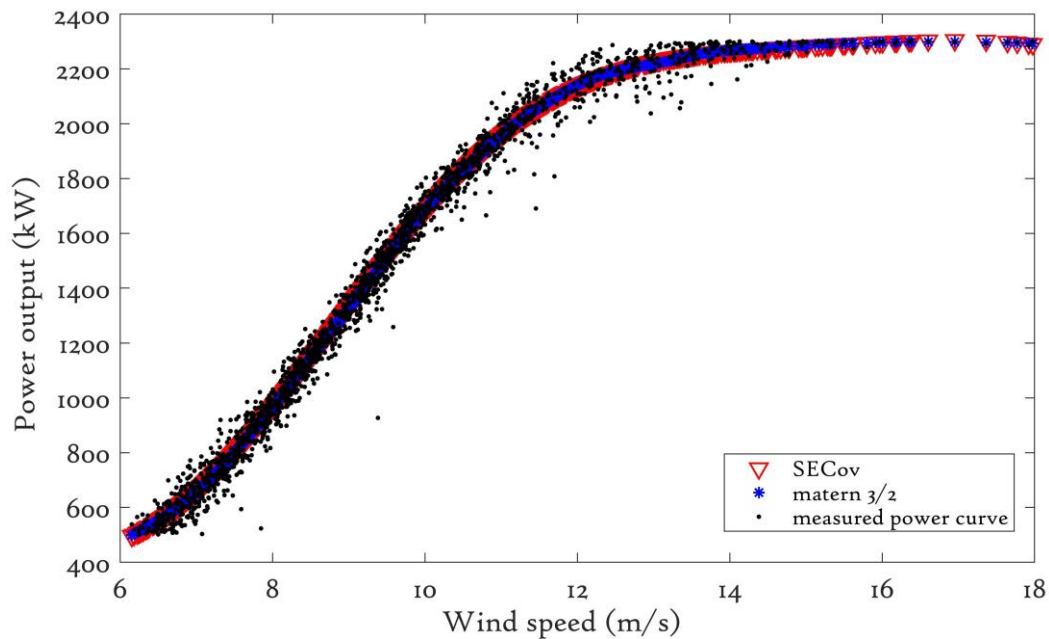


Figure 7: Estimated & measured power curve for SEcov and Matern 3/2

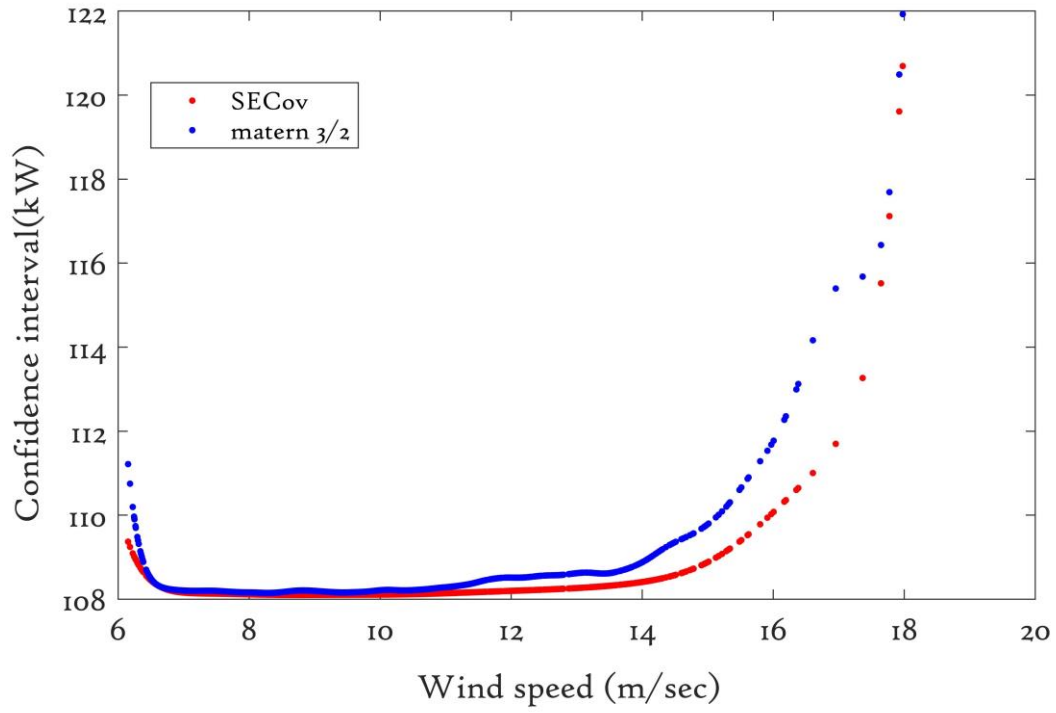


Figure 8: Uncertainty analysis using CI SECov and Matern 3/2

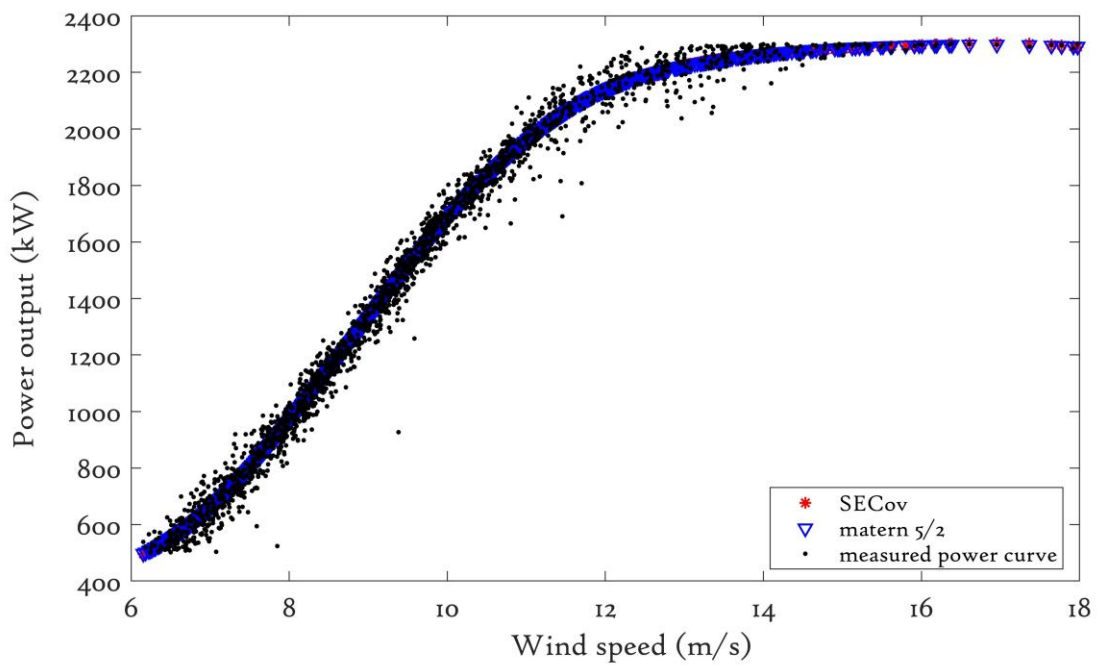


Figure 9: Estimated & measured power curve SECov and Matern 5/2

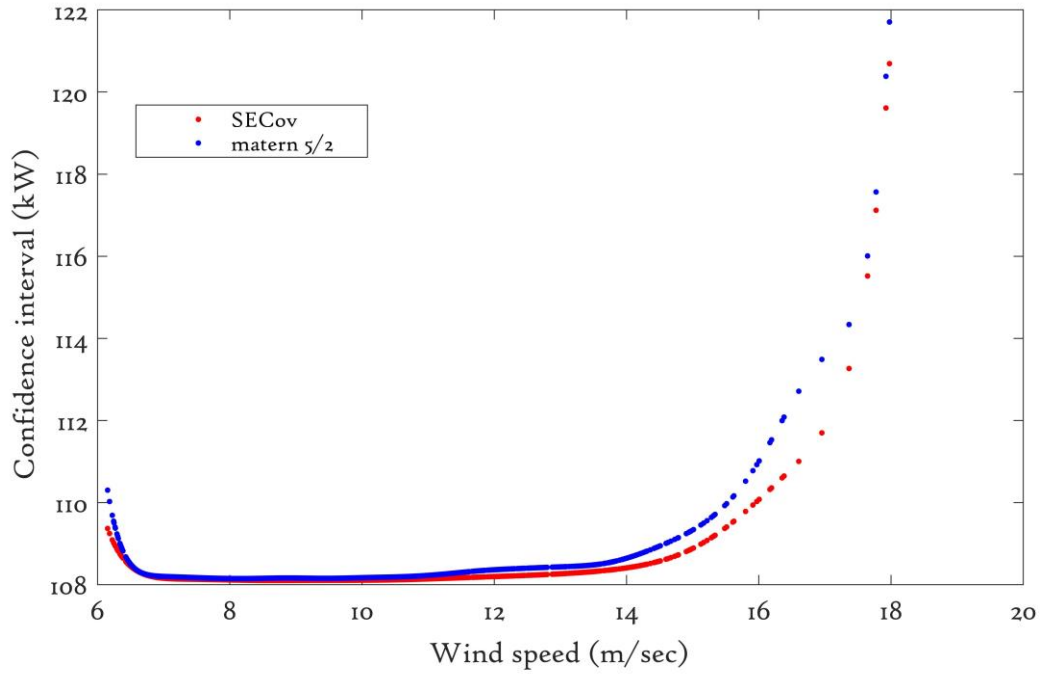


Figure 10: Uncertainty analysis using CI for SECov and Matern 5/2

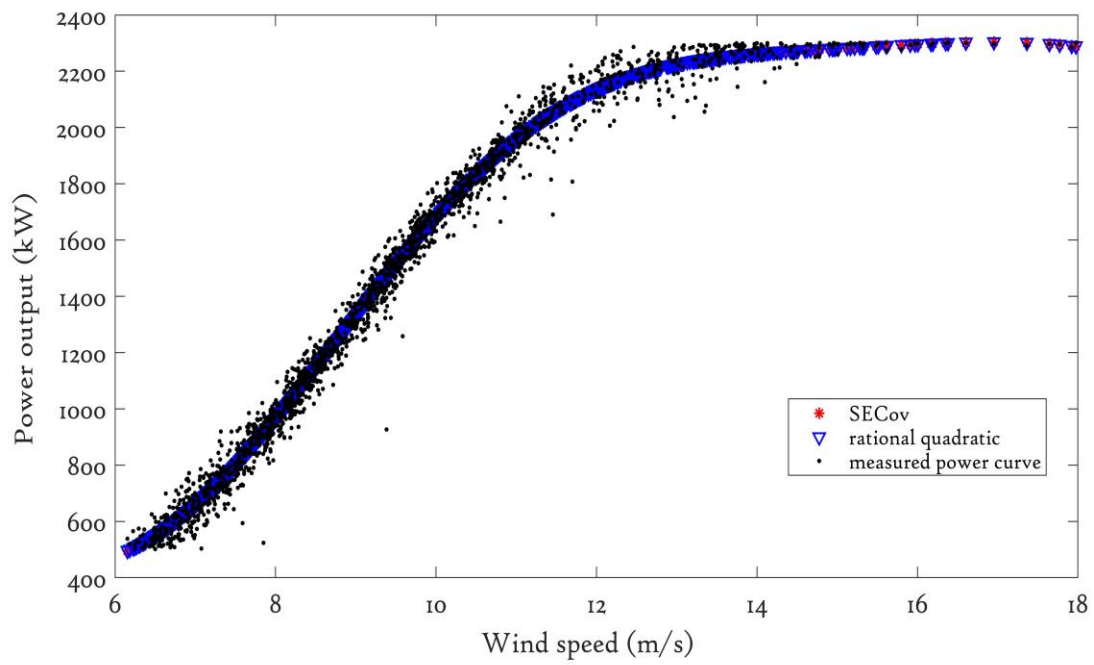


Figure 11: Estimated & measured power curve for SECov and RQ

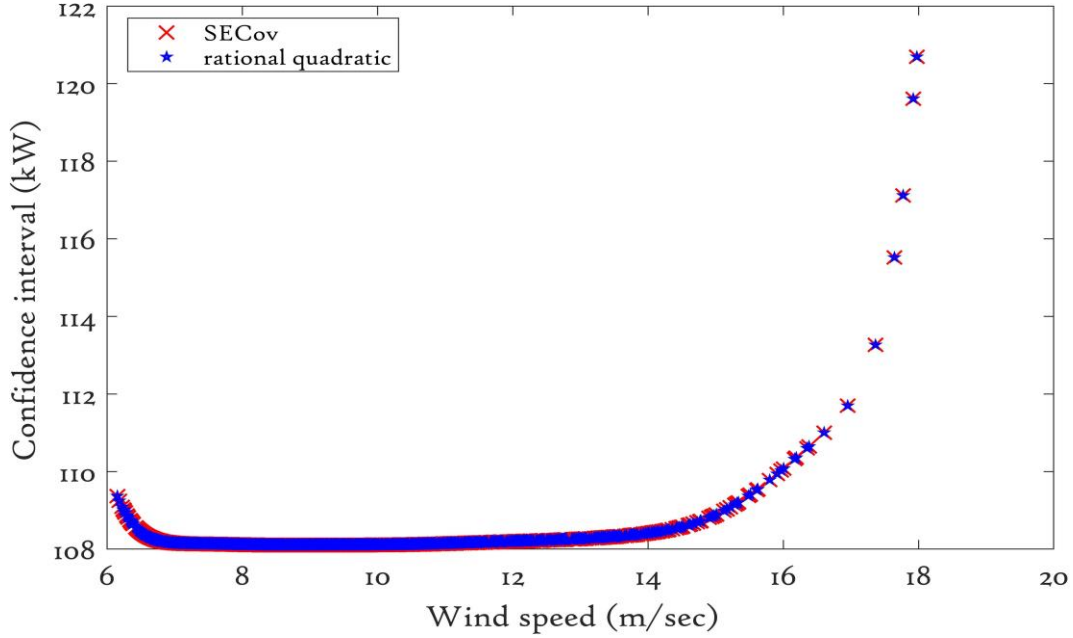


Figure 12: Uncertainty analysis using CI for SECov and RQ

## 8. Comparative Studies of Gaussian Process Models

In this section, a comparative study of the different stationary covariance functions and their influence on GP power curve fitting is presented. Figure 13 shows the power curve of the different GP models with respect to the measured power curve and no significant differences between the individual GP models can be observed. Model uncertainty is compared through confidence intervals as a function of wind speed, as shown in Figure 14. This suggests that the Matern 3/2 based GP model has the worst performance while both SECov and rational quadratic based GP models give the most accurate results. The GP based on Matern 5/2 performed better than Matern 3/2, but poorer than the SECov and rational quadratic GP models, especially in the region near to the rated wind speed. Well above rated wind speed very little difference can be discerned between the models, as shown in Figure 14.

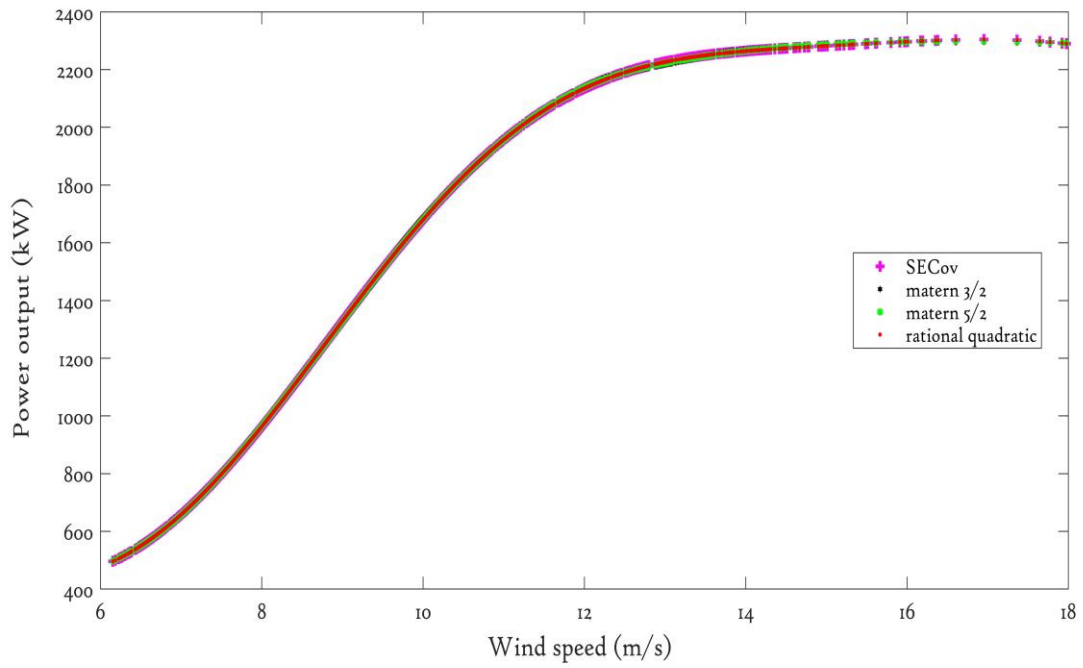


Figure 13: Comparative analysis of power curve fitting GP models

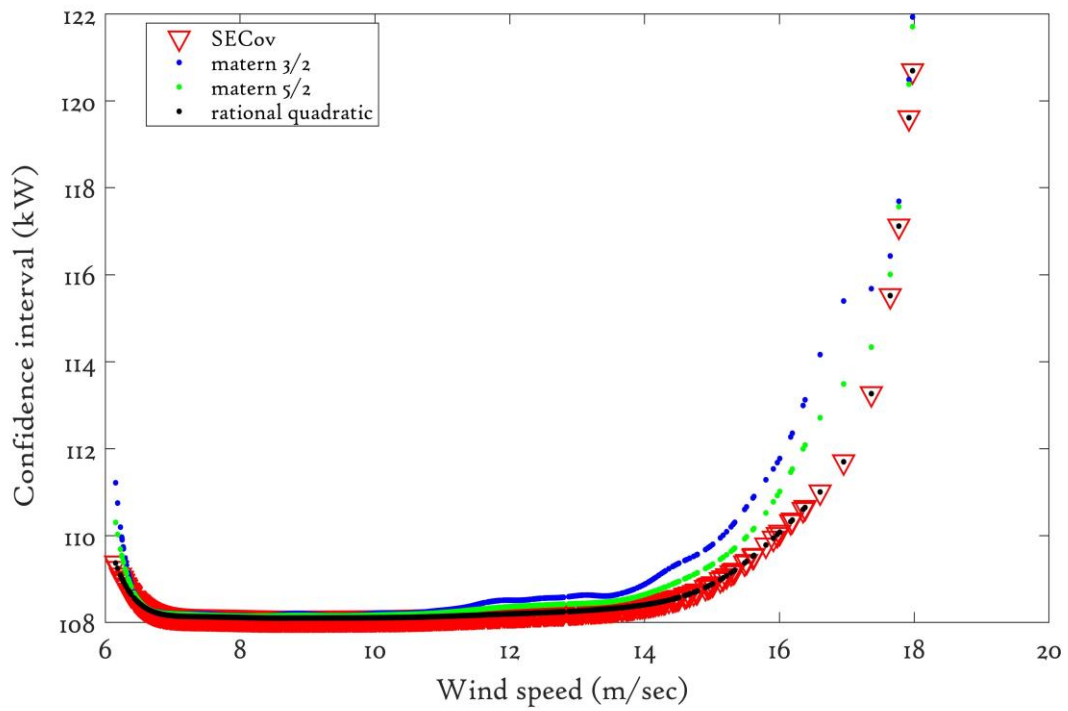


Figure 14: Comparative studies of GP models Uncertainty analysis using confidence intervals

The above conclusion can be confirmed via range of statistical performance error metrics such as MAE, RMSE and  $R^2$ . The difference between the measured and estimated values can be expressed

by mean absolute error (MAE): 
$$\text{MAE} = \frac{\sum_{i=1}^n \text{abs}(y_i' - y_i)}{n} \quad (25)$$

, or in terms of residuals, 
$$\text{MAE} = \frac{\sum_{i=1}^n (e)}{n} \quad (26)$$

To quantify the magnitude of the residuals (i.e. the difference between observed and modelled values, root mean square error (RMSE) is commonly used; defined as, [36]:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i' - y_i)^2}{n}} \quad (27)$$

where  $y'$  are the GP predicted values for  $n$  different predictions, and  $y$  are the measured values.

In terms of residuals this is: 
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (e)^2}{n}} \quad (28)$$

Another statistical measure, the coefficient of determination ( $R^2$ ), quantifies how close the data are to the fitted regression, [36], and calculated as the square of the correlation between predicted output and measured values (hence always in the range from 0 to 1 with values closer to 1 indicates better fitting of the model to the data). It is defined as  $R^2 = 1 - \frac{SSE}{TSS}$ ; where SSE is the sum of squared errors and TSS is the total sum of squares.

| MODELS                           | RMSE          | $R^2$         | MAE          | PREDICTION<br>SPEED       | TRAINING<br>TIME  | REMARKS                              |
|----------------------------------|---------------|---------------|--------------|---------------------------|-------------------|--------------------------------------|
| <b>GP SECov</b>                  | <b>55.017</b> | <b>0.9899</b> | <b>39.77</b> | <b>~26000<br/>obs/sec</b> | <b>45.502 sec</b> | <b>Strong smoothness,<br/>faster</b> |
| GP Matern 3/2                    | 54.834        | 0.9900        | 40.468       | ~16000<br>obs/sec         | 59.332 sec        | Poor smoothness                      |
| GP Matern 5/2                    | 54.956        | 0.9900        | 39.682       | ~17000<br>obs/sec         | 56.332 sec        | Less smooth compared to<br>SECov     |
| <b>GP Rational<br/>quadratic</b> | <b>55.016</b> | <b>0.9899</b> | <b>39.77</b> | <b>~20000<br/>obs/sec</b> | <b>143.57 sec</b> | <b>Strong smoothness,<br/>slower</b> |

Table 4: Statistical measure of each GP model based on covariance function

## 9. Conclusion and Discussion

Accurate estimation of WT output power, which is based on the power curve, is vital for wind farm feasibility assessments because the estimated energy output is a critical factor in cost analysis. When using GP models, the suitability of the covariance function is key to accurate power curve fitting.

The covariance between two given data points measures the degree to which the points are related. The selection and composition of covariance functions is a non-trivial task in GP modelling. In GP regression, both basis functions and their prior distribution are simultaneously described by choice of the covariance function. Despite having limitations such as in dealing with the whole of training data to perform estimation, [12], GP modelling is still considered as a highly effective modelling approach.

In this paper, an in-depth analysis of stationary covariance functions has been undertaken for the GP based power curve fitting. The results show that GP fitted power curve models based on rational quadratic covariance functions perform almost as well as the most commonly used SECov functions, but its prediction speed and time to train the model is slower, whereas GP models based on SECov have superior performance as shown in Table 4. In short, both covariance functions work pretty well in identifying a well fitted smooth function. The rational quadratic covariance function can be used instead of SECov for GP modelling if the data sets are not large and there is no limitation on training time and prediction speed. In contrast, the Matern 5/2 covariance function should only be used with caution, GP models based on Matern 3/2 perform poorly and are not recommended for wind turbine power curve modelling.

Future work involves using these concepts to identify a range of wind turbines failures.

## Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642108.



## References

- [1] World Wind Energy Association (WWEA) article on "world wind market has reached 486 GW from where 54 GW has been installed last year". Available online at <http://www.wwindea.org/11961-2/>; June 8, 2017.
- [2] Hameed Z, Hong YS, Cho YM, Ahn SH, Song CK. Condition monitoring and fault detection of wind turbines and related algorithms: a review. *Renewable & Sustainable Energy Reviews* 2009;13(1):1e39.
- [3] Lu B, Li YY, Wu X, Yang ZZ. A review of recent advances in wind turbine condition monitoring and fault diagnosis. In: *IEEE conference on power electronics and machines in wind applications (PEMWA'09)*, Lincoln; 24e26 June, 2009, pp. 1-7.
- [4] P. Tavner, *Offshore Wind Turbines Reliability, Availability and Maintenance*, The Institution of Engineering and Technology, 2012.
- [5] M. Lydia, S.S. Kumar, G.E.P. Kumar, Advanced algorithms for wind turbine power curve modeling, *IEEE Trans. Sustain. Energy* 4 (3) (2013), pp. 827-835.
- [6] S. Gill, B. Stephen, S. Galloway, Wind turbine condition assessment through power curve copula modeling, *IEEE Trans. Sustain. Energy* 3 (1) (2012), pp. 94-101.
- [7] Zdenko Simic and Vladimir Mikulicic, "Small wind off-grid system optimization regarding wind turbine power curve," *AFRICON 2007*, Windhoek, 2007, pp. 1-6. doi: [10.1109/AFRCON.2007.4401601](https://doi.org/10.1109/AFRCON.2007.4401601).
- [8] L.T. Qin, S.S. Liu, H.L. Liu, Y.H. Zhang, Support vector regression and least squares support vector regression for hormetic dose- response curves fitting, *Chemosphere* 78 (3) (2010): 327-334. doi: [10.1016/j.chemosphere.2009.10.029](https://doi.org/10.1016/j.chemosphere.2009.10.029).
- [9] N. Chen, Z. Qian, I. T. Nabney and X. Meng, "Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction," in *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 656-665, March 2014. doi: [10.1109/TPWRS.2013.2282366](https://doi.org/10.1109/TPWRS.2013.2282366).

- [10] Ravi Kumar Pandit, David Infield. SCADA-based wind turbine anomaly detection using Gaussian Process models for wind turbine condition monitoring purposes. IET Renewable Power Generation, vol. 12, no. 11, pp. 1249-1255, 2018 .doi: [10.1049/iet-rpg.2018.0156](https://doi.org/10.1049/iet-rpg.2018.0156).
- [11] Ravi Kumar Pandit, David Infield. Performance assessment of a wind turbine using SCADA based Gaussian Process model. International Journal of Prognostics and Health Management (IJPHM), vol. 9, no. 023, pp.8. doi: <https://www.phmsociety.org/node/2492>.
- [12] C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X.
- [13] Jin Zhou, Peng Guo and Xue-Ru Wang, "Modeling of wind turbine power curve based on Gaussian process," 2014 International Conference on Machine Learning and Cybernetics, Lanzhou, 2014, pp. 71-76. doi: 10.1109/ICMLC.2014.7009094.
- [14] M. Schlechtingen and I. F. Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," Mech. Syst. Signal Process., vol. 25, no. 5, pp. 1849–1875, 2011.
- [15] A.Kusiak, H. Zheng, and Z. Song, "Models for monitoring wind farm power," Renewable Energy, vol. 34, no. 3, pp. 583–590, 2009.
- [16] J.-Y. Park, J.-K. Lee, K.-Y. Oh, and J.-S. Lee, "Development of a novel power curve monitoring method for wind turbines and its field tests," IEEE Transactions on Energy Conversion, vol. 29, no. 1, pp. 119–128, 2014.
- [17] S. Mathew, Wind Energy-Fundamentals, Resource Analysis and Economics, Springer, Berlin, Germany, 2006.
- [18] Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, British Standard, IEC 61400-12-1, 2006.
- [19] Ping Li, Songcan Chen, A review on Gaussian Process Latent Variable Models, CAAI Transactions on Intelligence Technology (2016).
- [20] Neal, R. M. Bayesian Learning for Neural Networks. Springer, New York. Lecture Notes in Statistics, 118, 1996.

- [21] Online article on different types of covariance functions and its applications, available at <http://evelinag.com/Ariadne/covarianceFunctions.html>
- [22] Noel Cressie, Hsin-Cheng Huang. (2012). Theory and Methods Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. Journal of the American Statistical Association. Volume 94, 1999 - Issue 448. Pages 1330-1339.
- [23] Abhirup Datta, Sudipto Banerjee, Andrew O. Finley & Alan E. Gelfand. Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. Journal of the American Statistical Association Volume 111, 2016 - Issue 514. Pages 800-812.
- [24] Bowman, Frank Introduction to Bessel Functions (Dover: New York, 1958). ISBN 0-486-60462-4.
- [25] Abramowitz and Stegun. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. ISBN 0-486-61272-4.
- [26] The Kernel Cookbook: on Covariance functions by David Duvenaud Available online at <http://www.cs.toronto.edu/~duvenaud/cookbook/>
- [27] Andrew Gelman, Abdel H. El-Shaarawi and Walter W. Piegorsch, Prior distribution. Volume 3, pp 1634–1637, Encyclopedia of Environmetrics. (ISBN 0471 899976). John Wiley & Sons, Ltd, Chichester, 2002.
- [28] Pandit RK, Infield D, Carroll J. Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy. Wind Energy. 2018;1–14. <https://doi.org/10.1002/we.2285>.
- [29] Jasper Snoek, Hugo Larochelle, Ryan P. Adams.' Practical Bayesian Optimization of Machine Learning Algorithms'. Available online at: [arXiv:1206.2944v2](https://arxiv.org/abs/1206.2944)
- [30] Kevin Cremanns, Dirk Roos.' Deep Gaussian Covariance Network'. Available online at: [arXiv:1710.06202v2](https://arxiv.org/abs/1710.06202).
- [31] Gaussian Process Regression Models, Matlab toolbox version 2017b.
- [32] Gabriel Kronberger, Michael Kommenda.' Evolution of Covariance Functions for Gaussian Process Regression using Genetic Programming'. Available online at: [arXiv:1305.3794v2](https://arxiv.org/abs/1305.3794)

- [33] J. Hartikainen and S. Särkkä, "Kalman filtering and smoothing solutions to temporal Gaussian process regression models," 2010 IEEE International Workshop on Machine Learning for Signal Processing, Kittila, 2010, pp. 379-384. doi: 10.1109/MLSP.2010.5589113.
- [34] S. Sarkka, A. Solin, and J. Hartikainen, "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing," IEEE Signal Processing Magazine, vol. 30, no. 4, pp. 51–61, 2013.
- [35] Ravi Kumar Pandit, David Infield. Comparison of binned and Gaussian Process based power curves for condition monitoring purposes. Journal of Maintenance Engineering (vol:2), University of Manchester, 2017. ISBN no :978-1-912505-25-8.
- [36] Glantz, Stanton A.; Slinker, B. K. (1990). 'Primer of Applied Regression and Analysis of Variance.' McGraw-Hill. ISBN 0-07-023407-8.